

SW

PROPERTY OF
PRINCETON UNIVERSITY LIBRARY
RECEIVED OCT 23 1939

The PSYCHOLOGICAL RECORD

OCTOBER, 1939
Vol. III No. 16

SELF-DERIVED NORMS FOR INSTITUTIONS

GRACE H. KENT



THE PRINCIPIA PRESS, INC.
BLOOMINGTON, INDIANA

Price of this number, 30 cents

EDITOR
BUSINESS EDITOR

J. R. KANTOR, *Indiana University*
C. M. LOUITT, *Indiana University*

DEPARTMENTAL EDITORS:

ABNORMAL	EDMUND S. CONKLIN, <i>Indiana University</i>
CHILD	HELEN KOCH, <i>University of Chicago</i>
CLINICAL	B. M. CASTNER, <i>Yale University</i>
COMPARATIVE	E. A. CULLER, <i>University of Rochester</i>
EDUCATIONAL	J. G. PEATMAN, <i>City College of New York</i>
EXPERIMENTAL	B. F. SKINNER, <i>University of Minnesota</i>
MEDICAL	NORMAN CAMERON, M. D., <i>Cornell Medical College</i>
PHYSIOLOGICAL	C. F. SCOFIELD, <i>University of Buffalo</i>
PSYCHOMETRICS	J. P. GUILFORD, <i>University of Nebraska</i>
SOCIAL	NORMAN C. MEIER, <i>University of Iowa</i>
ASSISTANT EDITOR	J. W. CARTER, JR., <i>Indiana University</i>



The Principia Press, Inc., has undertaken the publication of this co-operative journal to afford authors the opportunity of immediate publication at the least possible expense. The present low cost of publication and possible future reductions, depend entirely upon the number of subscriptions. The subscription price has been made low to induce individuals to subscribe. Under the Articles of Incorporation of the Principia Press no profit can be made on any of its publications. Therefore an increase in the number of subscribers will be reflected in reduced costs to authors and in increase in the number of pages published annually.

EDITORIAL BOARD. The above named board of associate editors have agreed to serve in an active capacity for a period of three years, and all manuscripts will be submitted to them.

MANUSCRIPTS may be sent to the appropriate associate editor or to Dr. J. R. Kantor. Longer papers (30 or more typewritten pages) in the above mentioned fields are especially desired.

COSTS. The cost of publication will be \$2.00 per page of approximately 470 words. The charge for line cut illustrations will be \$2.00 regardless of size. Special charges will have to be made for complex tabular matter and for half-tone or colored illustrations.

REPRINTS. One hundred copies of each paper will be supplied gratis. Additional copies may be purchased in any quantity at a rate which will be set when the author is informed of the cost of publication.

ROYALTIES. Fifty per cent of the net income from the sale of individual copies, or from copies sold as part of back volumes, will be credited as royalties to the author's account. Royalties cannot be paid on income from subscriptions current in the year of publication.

SUBSCRIPTIONS. The subscription price is \$4.00 per year for an annual volume of approximately 500 pages. Individual papers will be sold at an advertised price, which will depend upon the size of the article. Foreign subscriptions will be \$4.50.

CORRESPONDENCE concerning editorial matters should be addressed to Dr. J. R. Kantor, *Indiana University*. Business correspondence should be addressed to Dr. C. M. Louttit, *Principia Press, Inc., Bloomington, Ind.*

SELF-DERIVED NORMS FOR INSTITUTIONS¹*

BY GRACE H. KENT

Danvers State Hospital, Hathorne, Massachusetts

I. INTRODUCTION

For certain classes of clinical subjects it is more important to know a person's relative standing in his own special group than to know how he compares with the community as a whole. Norms for mental tests are relative at best,—we have no absolute criteria for the entire population. Norms based upon a large and varied group of subjects offer a closer approach to a hypothetical absolute standard than is furnished by norms derived from a smaller group, but they are still relative.

The principle of relative criteria is widely recognized in the use of tests for classification of school children. Even when the tests are individually presented and when the results are expressed in terms of "mental age" or "IQ," it is primarily a child's rank in his own group that determines his placement in the upper, middle, or lower section of a grade which is to be divided into three approximately equal sections.

The standardization of mental tests at the adult level presents inherent difficulties as compared with standardization for children of school age, because there are no unselected adults to whom test standardizers have easy access in sufficient numbers. Public schools of different and varied communities offer a relatively representative group of children from whom norms may be obtained; but it is a very different matter to obtain a representative sampling of the population for persons outside of the age-range for compulsory

¹ The assistance of Dorothy C. McLeod was largely what made it seem possible to undertake this study. Miss McLeod was almost wholly responsible for the collection of data from foreigners up to 1934, and this material was to have been hers for analysis and publication had she remained in the service of the hospital. In later years valuable assistance has been received from Alice (Schoenfuss) Foster, Faith Kellogg, Josephine Tinsley and others.

* Recommended for publication by Dr. C. M. Louttit, September 8, 1939.

education. Not until the appearance of the Wechsler system (1) have we had any mental test properly standardized for subjects in middle life.

Even with the Wechsler tests and others which may be developed along similar lines, there is a place for such criteria as may be derived from a limited group, for the classification among themselves of the individuals comprising that group. Self-derived or autogenous norms are especially useful for what is in effect a self-perpetuating group—an institution which has a rapidly changing population but which retains its essential nature year after year.

A preliminary study of the possibilities of institutional test standardization was conducted by the writer in 1927 at the State Training School, Shirley, Massachusetts. This institution stands between the state school for juvenile offenders and the state reformatory which receives young men in later adolescence. The Shirley school is specifically for boys of the mid-adolescent period, having a lower age limit of 15 and an upper one of 18. This school furnished an ideal group for an experiment on the establishment of autogenous norms. The age differences were negligible, and yet the mental levels of the boys ranged from 8 to 14 years. The admission rate was such that it required only six months to collect test scores from 200 boys examined on admission.

The tests used in this study were from the series later published as the Kent-Shakow battery (2), in its second edition. The following year these tests were again revised, and the older forms soon became obsolete. The third edition was introduced into the school, and the institutional norms were of no further use.

II. MENTAL TESTS IN DANVERS STATE HOSPITAL

Since the year 1929 it has been customary in this hospital to make some use of psychometric tests in the routine examination of newly admitted patients under sixty years of age, not so much for exact ratings as for better understanding of the patients.

A state hospital is anything but a favorable place for a study of self-derived norms. The patients differ so widely in so many ways that it seems almost presumptuous to classify them according to a variable so comparatively unimportant as their achievement in mental tests. However, there is no reasonable doubt that some-

thing can be learned about many patients by means of this instrument; and inasmuch as the routine examinations furnished much of the data required for the study, it has seemed worth while to collect the material for statistical treatment.

It is naturally impossible to include all patients who fall within the age limit, but it is believed that the patients who are examined—roughly about sixty per cent of the new cases—constitute essentially an unselected group of the hospital population. There are patients who are too excited to give any attention to a test, and others who are in so profound a stupor that no response can be obtained. Some are omitted because of physical disability, and some because their behavior is so disorderly that it would be unsafe to bring them to the examining office. There is occasionally a patient who refuses all cooperation, in which case there is nothing to be done unless or until he yields to the persuasive powers of the examiner. However, these circumstances which govern the selection of patients for psychometric examination have no known bearing upon the test achievement of a patient who is examined or omitted; and therefore it is believed that the patients included for examination are unselected cases in so far as their measurable mental capacity is concerned.

Inside this group there is some selection of the patients for whom any particular test is used. Orally presented tests cannot appropriately be used for a deaf patient, nor written tests for an elderly person who is not provided with glasses. Sometimes all performance tests have to be omitted because the patient's hands are so shaky that he cannot manipulate the materials. All possible effort is made to adapt the examination to the individual patient, by using those tests which are best suited to his mental level, his command of English, his education, his occupational interests, and his personal tastes; but the examination is not necessarily restricted to those tests which are considered valid for the formal report. When material on a given test is being collected for statistical use, the standard of suitability is lowered appreciably in order to make the group of patients receiving this test as inclusive as possible. No appropriate test is omitted to make place for one which is less important to the individual examination; but an additional test may be included for the purpose of giving the hospital norms a wider range of applicability.

III. LANGUAGE TESTS FOR ENGLISH-SPEAKING PATIENTS

Seven tests, selected by a long process of elimination, have been found passably suitable for nearly all our psychotic subjects who can be given a psychometric examination of whatever kind. The oldest of these is Woodworth-Wells "Hard Directions" (3), published in 1911. This test was adopted by the writer in 1928 and was included for standardization in the Kent-Shakow battery (2), which furnished also five other tests of this series. The remaining one is the writer's "Emergency Test" (4). Information concerning the sources and standardization of these tests may be summarized diagrammatically:

Woodworth-Wells, 1911 Formulated at Worcester State Hospital, 1924; revised 1926 and 1928.	<div style="display: inline-block; vertical-align: middle;"> Hard Directions Information Similarity (vocabulary) Essential Property Essential Difference Arithmetical Reasoning Emergency Test (middle scale) </div>	<div style="display: inline-block; vertical-align: middle; font-size: 3em; margin: 0 10px;">}</div> Group standardized, without time limit, 1928-1933. Standardized 1930.
---	--	--

This battery, in whole or in part, has been used for a large majority of the literate patients examined since 1930; not excepting those of the lower levels for whom the Stanford-Binet scale is also used, nor those of the upper levels who receive a supplementary examination by Army Alpha.

By preference the seven tests are used as a battery rather than singly, but it is not invariably possible to obtain a complete record. There are patients who give fair cooperation in the early part of the examination but whose limit of good-natured compliance is reached before the completion of the series. Arithmetical Reasoning is the test which most frequently elicits a flat refusal. Emergency Test is the one most likely to be used singly, because it can be given to a subject who cannot see to read.

Each of these tests is comfortably discriminative for the mental levels from nine to fourteen years, a range which includes a very large proportion of the patients. But although the tests are very useful, the norms are highly unsatisfactory; both because the

standardization is crude and inadequate, and especially because any norms derived from school children are somewhat misleading as applied to adult subjects. A "mental age" of 10 or 12 years does not have the same meaning for a person in middle life as for a child or adolescent.

IV. CONSTRUCTION OF HOSPITAL NORMS

For each of these seven tests the scores obtained from hospital patients have been collected in a continuous series up to the number 1,000, for the establishment of self-derived criteria. The 1,000-score goal was reached for Emergency Test far ahead of the other tests; and Arithmetical Reasoning and Hard Directions were the last two to be completed. Any particular record which did not seem to represent satisfactory cooperation was thrown out at the time of the examination; but the satisfactory records contributed by the patient were retained, as it did not at any time seem possible to collect 1,000 complete battery records. There is naturally considerable overlapping of cases among the different tests, but so far as known the 1,000 cases are not identical for any two tests.

The six written tests were first standardized with no time limit other than what the children imposed upon themselves; but each test has also some very tentative norms based upon scores achieved in the first two minutes. The usual form of presentation is to hand the subject a red pencil at the end of two minutes, so as to obtain two scores. Little use is made of the timed scores for hospital patients, some of whom spend a half-hour on a two minute task. However, the timed scores were collected independently until there were 200 complete battery records scored both ways. The results were then compared by finding the median "mental age" ratings with and without the time limit. It was found that 71 per cent of the patients achieved higher ratings by the untimed work. For 17 per cent the advantage of the untimed achievement was 3 or 4 years, and the average advantage for the entire group was 1.2 years. Inasmuch as the measurement of pathological retardation is not the purpose of the psychometric examination, it appeared that the majority of psychotic patients can be scored more significantly on their untimed work. On the strength of this indication the recording of the timed scores was discontinued.

It required six years, 1931-1937, to bring the series to completion. The original test records were not preserved, as they were so bulky as to involve undue fire risk. At the end of each month the scores were copied in condensed form, each set of scores accompanied by the patient's age and identifying number, after which the originals were destroyed.

The scores for each of the seven tests have been divided into

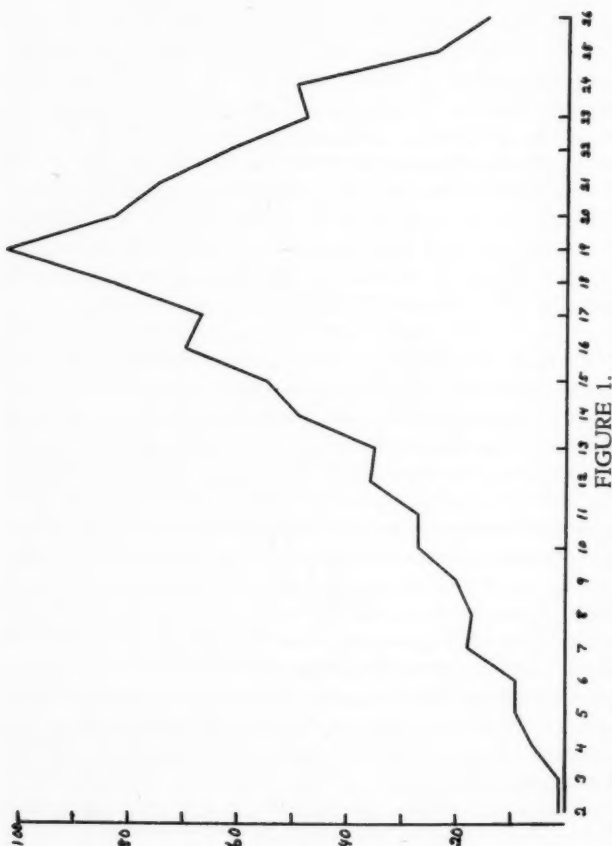


FIGURE 1.
Test *Essential Property*, having a perfect score of 26. Distribution of scores obtained from 1,000 state hospital patients. Vertical lines at base indicate the scores in which the decile divisions fall. Median score 18.

deciles, and the rating by these tables is used as a check on the customary evaluation of the test scores. In addition to the "mental age" by this battery, it is reported that a patient's scores place him in the upper thirty per cent or the lower ten per cent or the middle twenty per cent as compared with 1,000 hospital patients previously examined. The term "decile" is not used in the formal report.

The distribution of scores is illustrated graphically for Essential Property, selected as being the test which yields the most typical distribution (Figure 1). Arbitrary division into deciles was adopted as a means of reducing the results of the seven tests to a common denominator. It is not known to what extent the different tests are comparable, nor which of them is most truly representative of the hospital population; but it is a safe guess that a rating based on the median of several tests is more trustworthy than a rating by any single test. To this end it was necessary that the results should be treated uniformly, without regard to the natural distribution.

No analysis has been made on the basis of sex, because it appeared from preliminary studies that sex difference in achievement was negligible. It was assumed that age differences would be more significant, and it seemed entirely possible that 200-300 cases at each of four age groups would have greater value for norms than 1,000 cases of mixed ages. Tests based so largely upon school tasks might reasonably be expected to show higher scores for subjects in their twenties than for those in their fifties. However, the results of such analyses as have been made do not support this assumption. The scores for Hard Directions and for Essential Property have been tabulated separately for four age groups by decades, with the results shown in Table I.

TABLE I.
Age Differences in Scores

Age-group	Median score	Median score
	Hard Directions	Essential Property
under 30	28	18
30-39	30	19
40-49	28	18
50-59	29	18

The number of cases in each group differs for the two tests. For each test it is over 300 for the youngest group and under 200 for the oldest one.

Two performance tests have been standardized for the institution, and other studies of like nature are in progress. But it is a slow process to collect sufficient material for statistical treatment, because these tests require separate standardization for the sexes. All performance tests which have been used regularly for both sexes show higher scores for men than for women. Age is also a factor which cannot be disregarded. The color cube test (6), which has been used chiefly for women, appears to show a slight but consistent falling off with advancing age. The present indications are that both age and sex are more significant for performance tests than for language tests.

V. PERFORMANCE TESTS FOR FOREIGN SUBJECTS

No mental tests with which the writer is familiar are even passably well adapted to the illiterate immigrant who speaks little or no English. Manipulative tests which call for no use of language are occasionally useful, but only within narrow limits. Any test requiring the use of a pencil is of necessity unfair to the adult subject who has not been taught to write. Any test involving well-coordinated handling of small blocks is unfair to the peasant whose work in the fields has developed his larger muscles to the neglect of the smaller ones. Tiny pictures are not readily understood by the immigrant laborer, especially when the pictured objects are strange to him. Such a test as Myers Pantomime (5), specifically intended to be applicable to subjects of any language or environment, requires fine ocular adjustments to which the illiterate person is wholly unaccustomed. When the subject's clumsy performance of tests such as these is scored by speed and evaluated by comparison with the achievement of American children who from early infancy have played with blocks and pictures and pencils, the injustice is obvious.

The blundering use that has been made of standardized tests in the study of immigrants may be illustrated by the case of a Jewish woman of forgotten identity who was sent to this hospital some ten years ago under the diagnosis "Psychosis with Mental Defi-

ciency." On what tests this diagnosis was based is not known, but the patient's test achievement in this hospital is remembered as being uniformly low. Elated and over-talkative, she was quite unable to give much attention to any task or question, frequently responding: "My boy, he could tell you." Her explanation of her condition was that she became "*too happy*" when informed that her ten-year-old son had received a double promotion in school. During the interview she talked almost incessantly about the boy, abundantly able to describe her feelings to the examiner although her English was understood only in broken phrases.

This patient was undoubtedly psychotic, and therefore the diagnosis was of minor importance. A more serious problem is presented when there is question concerning the deportability or the criminal responsibility of a person who is not demonstrably psychotic; or when the issue involves a mother's fitness to have the legal custody of her children. It is when there is some matter of far-reaching social importance at stake that our aid is most eagerly sought, and it must be acknowledged that standardized tests contribute very little to the solution.

Theoretically, tests for immigrants should be devised by persons familiar with the occupations and living conditions in the mother-country of the people for whom the tests are to be used. Presumably Ellis Island is the only place in this country which offers in sufficient quantity the material required for standardization, and it is by no means certain that the material obtained there is qualitatively satisfactory. But if adequate standardization of tests for foreigners appears to be almost a hopeless undertaking, it is still possible to make some use of such tests as we have. To this end it has seemed worth while to attempt the development of special norms, derived from immigrants whose command of English is insufficient for the use of language tests.

Over 300 immigrants admitted to this hospital since 1932 have been observed by means of whatever non-language tests have been available. This group of patients does not include foreign-born persons who were brought here in early life and who have grown up in this country. It is a group of old-world persons, nearly all in middle life.

This hospital is at a disadvantage in that most of the perform-

ance test equipment has been acquired or designed to meet the needs of the traveling clinic. Test materials used for immigrants should be large enough to fit a man's hand, as different as possible from the apparatus used. Some liberties have been taken with the presentation and scoring of tests, especially with reference to time limits. If the study were being started at this time it would seem advisable to go much farther in adapting each test to the ability of the typical immigrant, without regard to standard procedure of presentation. There has been no deviation from the method adopted in 1932 for the tests then in use; but some of the tests introduced more recently have been used wholly without reference to their norms.

Records obtained from male patients numbered 200 before those of the female patients reached the number 100. This is due largely to the fact that it is easier to obtain cooperation from men than from women—a fact which holds for native population as well as for immigrants.

The only test which has been used strictly in accordance with published instructions for presentation and scoring is the Kohs Block Design test as modified by the writer (6). This test is considered more suitable for women and children than for men, as the blocks are too small to be handled comfortably by a man.

TABLE II.

SCORES BY KENT-KOHS COLOR CUBE TEST, DIAGONAL SERIES

Decile	Immigrants 200 men	Immigrants 100 women	Americans 1,000 women
I.	2	0	2
II.	5	2	5
III.	8	5	8
IV.	10	8	11
V.	13	10	16
VI.	16	13	22
VII.	23	17	29
VIII.	29	25	38
IX.	38	36	49
X.	54	52	68

However, in this as in all other performance tests, the scores of the men are uniformly higher than those of the women. The two sets of scores are presented in Table II, with the scores of 1,000 American women included for comparison. It will be noted that the American women—except at the lower levels—achieved higher scores than those of the immigrant men.

This institution is evidently not large enough to furnish the material needed for this study. No detailed analysis can be made of so small a number of cases. The development of norms applicable to institutionalized immigrants is a project for collaborative effort among several institutions.

VI. MENTAL TESTS FOR HOSPITAL EMPLOYEES

This study is less than a year old and there are as yet no results to be reported; but it is so closely allied to the study of test records obtained from hospital patients that it seems in order to mention it here. Somewhere nearly all the mistakes that could easily be made in collecting and assembling the data had already been made in the course of the older project, and it was possible to take advantage of them in planning a system of tests for an entirely new group.

State hospital attendants constitute a selected group which is not entirely constant from year to year. In times of wide-spread unemployment persons of higher mental level are found among the applicants than would be available for this work in more prosperous times. This may apply equally to men and women, but there are other selective factors which may affect either group independently. Whatever their comparative test achievement, they should be treated as two separate groups rather than one group. Already it has been observed in this hospital that the turnover is much heavier for women than for men. Whether this is local or general is a matter of conjecture.

At present an individual examination occupying at least one hour is being given to each new attendant as his first task. Twelve written tests are regularly used, including the six written tests referred to in section III. The others are four tests from the Kuhlmann-Anderson series (7) and two unstandardized tests which are introduced experimentally. Eight of the twelve tests are given both with and without time limit, thus yielding two scores each.

All these tests are being specially standardized for the two groups of subjects. The scores are recorded on foolscap sheets (separate sheets for men and women), with the scores of a given subject in a horizontal line and the scores for a given test in a vertical column.

Performance tests are used in every examination, primarily to give the examiner an opportunity to observe the subject's reaction to a strange and unexpected situation. Several tests are being tried out, but as yet no particular test has been adopted for routine use.

This is a much better project for a state-wide institutional system than for a single institution. Conditions of employment in different hospitals of the same state have enough in common to attract applicants who have something in common. The employees whom an institution retains in service may be affected by local conditions which are peculiar to that institution, but the larger group of those who seek institutional employment may be treated as one class for test standardization. If a battery of tests could be agreed upon for the use of several institutions, it would require only a short time to collect data for norms which would be more useful than norms derived from children.

SUMMARY

For selected groups of subjects who have passed the developmental period, self-derived norms offer a criterion which is useful for the evaluation and interpretation of mental tests. The actual results upon which this study is based are primarily of local significance and are therefore not reported in full; but the method is one that can be recommended for use in other institutions.

For each of seven language tests the scores obtained from 1,000 state hospital patients have been arranged in decile groups, to aid in evaluating a patient's test achievement by comparison with his own group. Several performance tests are being similarly standardized for this institution.

No significant sex difference has been observed in the patients' reactions to language tests; but performance test records, from whatever group of subjects, show such sex differences as to indicate separate norms for men and women. Other things being equal, the achievement of the men is appreciably higher than that of the women.

Age differences should be recognized in classifying test material of whatever type. It appears, however, that they are more significant for performance tests than for language tests.

Several non-language tests have been partially standardized for the immigrant population of this institution. The subjects have been grouped according to sex, but not according to racial or national groups. The number of cases is not sufficient for the establishment of norms; but the results clearly indicate that the performance test as standardized upon American children is grossly unfair to the immigrant.

The method is especially recommended for state training schools in which tests are used as a routine measure. In such an institution the accumulated test records can be used for constructing a system, by means of which any subject may be rated as a member of his own group. Any desired test, standardized or not, may be gradually introduced into the system. It is possible also to weed out the less satisfactory items of a standardized test.

Self-derived criteria would be of special significance in a reformatory; or in any institution which has a rapid turnover of inmates while retaining its essential character as an institution.

State-wide collaboration is desirable for collecting data from a group which is too small to furnish material for its own norms. This is recommended for foreigners, for state hospital attendants, and for any other relatively homogeneous group which is scattered among various institutions.

NOTE: Sample mimeographed forms will be furnished on application, as long as the supply lasts, to any student interested in giving this plan a trial.

REFERENCES

1. Wechsler, David. *The Measurement of Adult Intelligence*. Williams and Wilkins, 1939.
2. Kent, G. H. Written Tests for the Clinic. *J. Genet. Psychol.*, 1934, 44, 49-68.
3. Woodworth, R. S., and Wells, F. L. Association tests. *Psychol. Monog.*, 1911, No. 57.
4. Kent, G. H. Oral Test for Emergency Use in Clinics. *Ment. Measur. Monog.*, 1932, No. 9.

5. Myers, G. C. *A Pantomime Group Intelligence Test*. Newson and Company, New York, 1922.
6. Kent, G. H. Modification of the Kohs Block Design Test. *J. Appl. Psychol.*, 1934, 18.
7. Kuhlmann, F., and Anderson, R. G. *Kuhlmann-Anderson Intelligence Tests*. Educational Test Bureau, Inc., Minneapolis.

